INSTITUT PASTEUR

Research in Microbiology 159 (2008) 67–73

www.elsevier.com/locate/resmic

Tracking microbial biodiversity through molecular and genomic ecology

Abstract

Molecular ecology and metagenomics applied to the study of microbial biodiversity are changing our comprehension of the biosphere. An impressive diversity of archaea, bacteria and, more recently, protists has been uncovered by molecular tools. Efforts to couple function to the phylogenetic diversity observed in natural environments are leading to the discovery of novel metabolisms and to a re-evaluation of the global ecological impact of known ones. Interesting questions relating to mechanisms of speciation and evolutionary trends at the smallest and largest phylogenetic scales are emerging.

Keywords: Molecular ecology; Genomic ecology; Metagenomics; Microbial biodiversity

1. Introduction

One of the first and most successful applications of molecular phylogeny was the recognition of the Archaea and the building of a tripartite tree of life by C.R. Woese and collaborators from the late 1970s. Since then, microbiology lives under a permanent revolution, being one of the most fast-moving scientific disciplines. Vast fields of exploration are opening up for microbiology, and exciting questions whose answers were thus far largely inaccessible by only classical approaches can now be tackled by their combination with molecular and genomic tools.

How many microbial species are there? Over the last two decades, the widespread use of molecular methods based on the amplification, cloning and sequencing of small subunit ribosomal RNA (SSU rRNA) genes from the environment led us to open the black box of prokaryotic diversity in natural communities. Operational taxonomic units (OTUs) based on SSU rRNA sequence similarity have become indispensable proxies to the otherwise intangible concept of microbial species. What is the real extent of this diversity? Despite many thousands of environmental gene and (meta)genome sequences accumulating in databases, not even an approximate answer is in close view, and the application of the required ecological and statistical tools that could help to provide fair estimates is just starting. To the colossal diversity of archaea and bacteria detected by molecular tools can be added a wealth of microbial eukaryotic sequences, which are revealing an unsuspected variety of protist taxa that had escaped identification by classical protistology. Although they fall out of our scope here, viruses on their own represent a whole understudied world. DNA and RNA viruses outnumber by and large prokaryotes in oceans and soils, and their diversity is only beginning to be roughly drawn by metagenomic analyses.

What does all that microbial biodiversity do? How do ecosystems ultimately work? Despite massive advances in evaluating microbial diversity, its actual contribution to ecosystem functioning is still elusive. Substantial progress is being made though improved methodologies, including multidisciplinary approaches and metagenomic analyses, in the context of more discovery-prone conceptual frameworks, which are partly made possible by the finding of novel, sometimes unexpected metabolisms that are changing our understanding of global biogeochemical cycles.

How do species evolve? Why are there so many lineages? How did major taxa split and diversify? The study of microbial evolution is also largely benefiting from environmental sequencing by allowing access to reservoirs of genotypes from uncultured organisms, thus providing raw data to study micro- and macroevolutionary issues.

In the following, we very briefly summarize current lines and perspectives of research about these questions.

2. The extent of microbial diversity

The intensive application of molecular techniques to describe microbial diversity in natural environments is yielding

0923-2508/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2007.11.019

a massive amount of data that reveals interesting common points and differences between prokaryotic and eukaryotic diversity parameters.

2.1. Prokaryotes

In most environments, SSU rDNA surveys have unveiled a diversity of prokaryotes that is orders of magnitude larger than ever thought. Comparisons between classical culture-dependent and molecular methods have revealed huge gaps and biases in our appreciation of microbial diversity. Only a tiny fraction, about 1%, of the actual prokaryotic diversity appears to be amenable to culture [19], but these estimates evolve with each methodological improvement. Soils offer a good example. Several hundred bacterial species can be isolated from soils, a few thousand of different sequence types (phylotypes or OTUs, usually defined on the basis of 97% sequence identity) retrieved by SSU rDNA surveys and more than 25,000 phylotypes by direct massive sequencing [34]. Moreover, re-evaluation of classical DNA reassociation data with new analytical methods provides estimates of several million species [13]. This estimate's dependence on the methods (and underlying species concepts in some cases) makes any trial to quantify global microbial diversity too premature. In addition, accumulation curves of SSU rDNA surveys are usually far from saturation, suggesting that a far larger number of sequences would be required to fairly represent natural communities. However, interesting efforts to borrow parametric and mostly non-parametric estimators of species/OTU richness from classical ecology are arising both for microbial prokaryotes and eukaryotes, even though their present reliability seems limited [8].

Many phylotypes appear to define novel high-ranked lineages, namely, new phyla intermixed in phylogenetic trees with those having cultured members. However, SSU rDNA-based phylogenies are often characterized by the lack of resolution of the deepest nodes, which can partly explain the discrepancies in the number of prokaryotic phyla estimated by different authors. A phylotype can be misplaced in trees because of insufficient phylogenetic signal and other tree reconstruction artifacts (deficient taxonomic sampling, mutational saturation, long branch attraction), so that it can be erroneously considered a ''novel phylum''. The case of the archaeon *Nanoarchaeum equitans* illustrates this point. SSU rDNA phylogenies suggest that it defines a new archaeal kingdom, but multi-marker phylogenies have shown that it is actually a fast-evolving member of the Thermococcales, misplaced in SSU rDNA trees by long-branch attraction [4]. This is probably the case for many candidate ''novel phyla''. Among current classifications, the number of prokaryotic phyla varies from 50 to 88, reflecting not only phylogenetic artifacts but also opposite taxonomic practices. Like zoologists and botanists several decades ago, microbial taxonomists currently range between two extremes: ''splitters'', who define new phyla for any divergent phylotype, and ''lumpers'', who make large units to keep as stable a taxonomy as possible. A compromise between the two is likely more realistic: several prokaryotic phylotypes cannot be ascribed to any known group and therefore define novel phyla, but many others can, and should not be used to artificially inflate the real number of phyla. Recent initiatives, such as Greengenes (http://greengenes.lbl.gov), allow a very useful cross-comparison of different taxonomic schemes. Fig. 1 shows only the 53 bacterial phyla that are acknowledged by at least 3 of the 5 taxonomic frameworks used in Greengenes, and 17 major archaeal lineages.

2.2. Microbial eukaryotes come into play

Compared with prokaryotes, molecular inspection of the diversity of microbial eukaryotes (protists) is still in its infancy, but it already shows contrasting tendencies to those seen in prokaryotes. Environmental protist phylotypes usually differ from known species sequences, but only very few of them define potential novel divergent lineages [27]. This suggests that the traditional protist description has been much more exhaustive than the prokaryotic one, something easy to explain taking into account the relatively large size and complex morphology of many protists. However, molecular approaches are decisive in demonstrating that not only are the protists of ''traditional'' size (>5 μm) very diverse, but also small nano- and pico-eukaryotes (<5 μm). Though understudied by traditional methods, small protists turn out to be the most speciose in many SSU rDNA molecular surveys [27].

Up to now, the majority of these surveys have been focused on marine planktonic communities, although several other ecosystems, such as freshwater, sediments and extreme environments, have also been studied [8]. In marine plankton, the largest proportion of protist phylotypes belongs either to the Heterokonta or the Alveolata (Fig. 1). Heterotrophic heterokonts dominate in surface waters together with photosynthetic picoalgae, whereas two groups of unidentified alveolates, Marine Alveolate Groups I and II, dominate in deep, aphotic waters [27]. Group II alveolates were soon recognized as relatives of the genus *Amoebophrya* and, very recently, species of the genus *Duboscquella* have been shown to branch within Group I [17]. Both *Amoebophrya* and *Duboscquella* belong to the Syndiniales, a poorly known group of parasitic dinoflagellates. Parasitism may therefore be extremely common in marine protist planktonic communities, most likely playing an unexpected major role in population control.

3. Microbial biodiversity and ecosystem functioning

To understand how microbial and, ultimately, all ecosystems work, identifying the different components of the community is but a very preliminary step that needs to be complemented with data about their functions (metabolism, lifestyle), interactions, spatial and temporal dynamics and environmental parameters. Developing ecosystem models that accommodate all this information is fundamental, but this is a long-term objective, as primary data connecting most microbes to their functions are still missing.

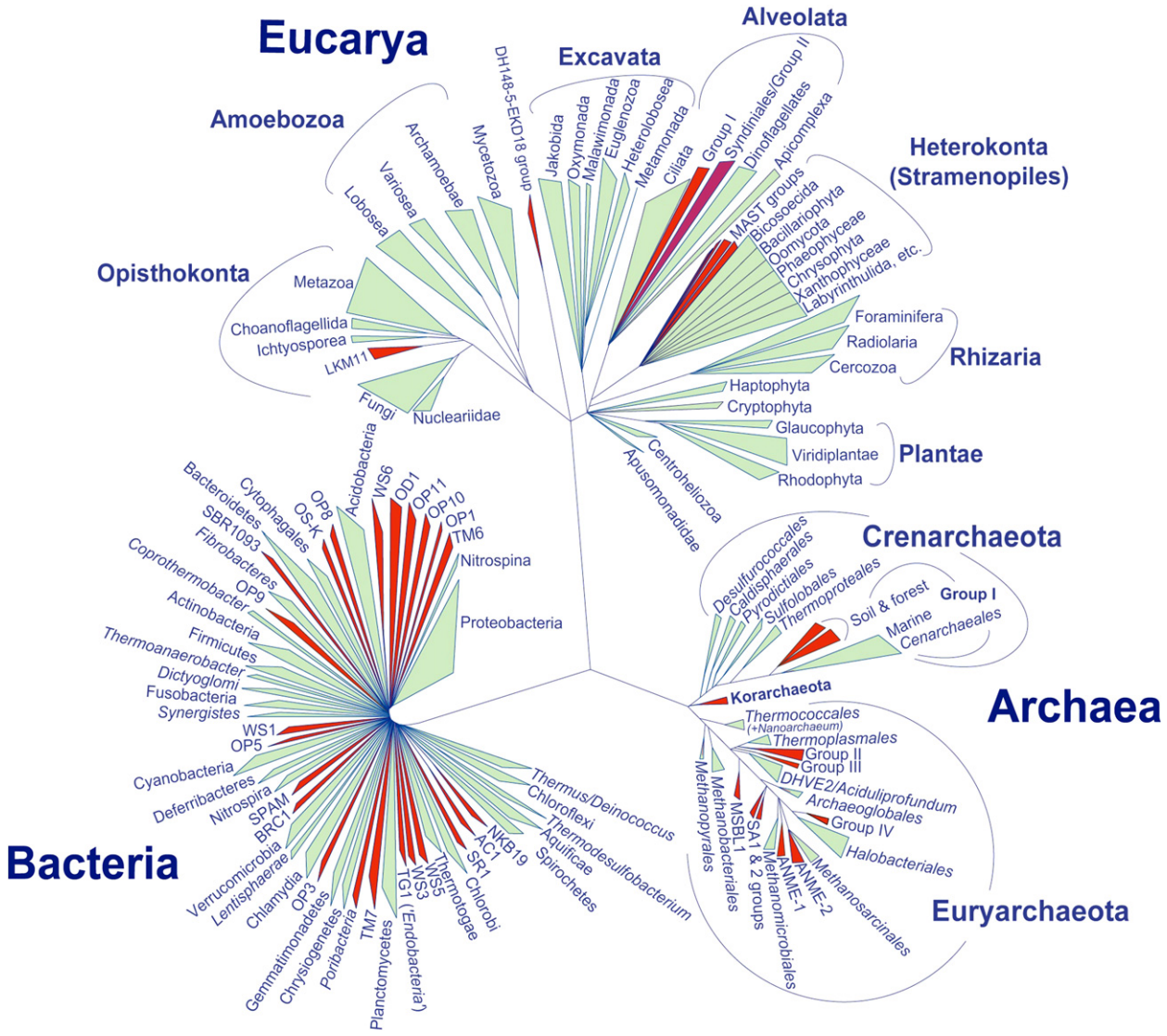

Fig. 1. Schematic phylogenetic tree of life based on current molecular knowledge (SSU rRNA and other molecular evidence). Green/light triangles represent phyla, divisions or groups of high taxonomic rank for which at least one member has been cultivated and/or properly described (e.g. many protist species); red/dark triangles represent candidate divisions or highly divergent lineages without cultivated/described species. The tree is highly simplified, only a fraction of known eukaryotic phyla are depicted, and, in the case of bacteria, only phyla and candidate divisions agreed upon by at least three classification systems, as acknowledged in Greengenes (http://greengenes.lbl.gov).

3.1. Linking molecular diversity and organism function

We ignore what most environmental lineages, identified exclusively by sequence, are doing in nature. Characterizing at best environmental physico-chemical parameters may help to formulate hypotheses about the metabolisms operating in a given sample. A variety of techniques enables visualizing individual microorganisms, linking them to particular phylogenetic groups, gathering information about their constituents and testing their activities in situ. Fluorescent in situ hybridization (FISH) is used to identify and quantify microbes from particular lineages with the help of fluorescent or confocal microscopy. Biomarker (lipid) analyses also provide information about the taxonomy and relative abundance of archaea and bacteria, often at subphylum levels. FISH can be coupled to microautoradiography (FISH-MAR) experiments to see whether targeted organisms metabolize particular labeled substrates added to the environment [42]. Simultaneous FISH and gold nanoparticle labeling enables visualizing cells belonging to specific groups, also by scanning and transmission electron microscopy, which can reveal particular ultrastructural features, and coupling X-ray diffraction analyses, the nature of associated mineral phases [14]. Stable isotope probing (SIP) enables tracking carbon flow [7] and that of other elements such as nitrogen and potentially sulfur or iron. Recently, Raman-FISH combining stable-isotope Raman spectroscopy and FISH has been applied to couple single cell identification and functional observations [18]. High-throughput methods leading to the simultaneous detection of phylogenetic and functional markers in single cells are actively being developed. Examples are multiplex microfluidic digital PCR on natural communities [30] or high-speed fluorescent activated cell

sorting followed by multiple displacement whole genome amplification and PCR screening [38].

Combining several of these strategies may prove successful. Noteworthy examples are the discovery of anaerobic oxidation of methane (AOM) by consortia of putative methanotrophic ANME archaea and sulfate-reducing bacteria [3,29] and that of anaerobic oxidation of ammonium to N_2 with nitrite (anammox) by abundant planctomycetes bacteria in the Black Sea [23]. Despite this, cultivation is the most direct means to associate functions to new microbes and to test physiologically potential functions eventually deduced from gene amplification or metagenomic data. In turn, isolation offers complete genomes that can be linked to metabolic functions and used as a reference for metagenomic annotation and prediction. Notorious efforts have been made in recent years to isolate representatives of uncultivated groups that appeared to be quantitatively important and potentially key players in environmental processes. Among the most famous are candidates *Pelagibacter ubique* and *Nitrosopumilus maritimus*, which waited many years before they could be cultivated. *P. ubique* is a representative of the widespread marine gammaproteobacterial SAR11 group. It is a strict oligotroph accounting for from 25% to occasionally 50% of bacterioplankton cells in surface waters, and possesses the smallest genome known for a free-living heterotroph (1.3 Mb) [15]. Most marine planktonic bacteria are likely oligotrophic, but common cultivation media are exceedingly nutrient-rich and favor opportunistic growth, a fact that illustrates how far laboratory cultivation conditions are from those in natural systems. Fortunately, better knowledge about the latter combined with high-throughput cultivation techniques will overcome part of these problems, as exemplified by the isolation, after careful mimicking of its natural habitat, of one member of the prevalent deep-sea vent euryarchaeotal group DHVE2, *Aciduliprofundum boonei*, an obligate thermoacidophilic sulfur- and iron-reducing heterotroph that can make up to 15% of the archaeal population [33]. *N. maritimus* is the first cultivated representative of the marine Group I crenarchaeota, another abundant planktonic group accounting for up to one third of prokaryotic cells in deep oceanic waters [21]. It grows by aerobically oxidizing ammonium to nitrite, a metabolic ability so far unknown for the archaea [22]. Other metabolisms unexpected in known groups have been reported, suggesting that some metabolic capabilities may be easily transferred horizontally. For instance, the recently isolated candidatus *Chloracidobacterium thermophilum*, is the first member of the Acidobacteria having chlorophyll genes and performing aerobic photosystem-based phototrophy [5]).

3.2. Massive sequencing approaches

Predicting functions from whole community genomes (metagenomes) is a powerful approach. However, metagenomic analyses, which ultimately rely on comparative genomics of cultivated organisms, is not exempt from difficulties, the most important being exponentially amplified annotation errors in databases that may lead to inappropriate predictions and huge amounts of accumulating sequence data that urgently call for development of better adapted bioinformatic tools. Despite this, metagenomics has led to interesting functional discoveries, and its biotechnological and drug discovery potential is being exploited by industrial companies. Cloning and sequencing of large genome fragments containing phylogenetic markers enable coupling phylogenetic groups to particular functional genes. Paradigmatic examples were the discovery of rhodopsin-based phototrophy in marine proteobacteria [1], a wide range of aerobic anoxygenic phototrophic planktonic bacteria [2] and the suspicion, based on the presence of *amo* genes, that group I crenarchaeota might oxidize ammonium [36]. Massive random shotgun sequencing of high diversity environments such as oceans and soils permits the identification of the most frequently represented functional genes in metagenomes and points to metabolisms that are relevant in a given ecosystem. Shotgun sequencing of the Sargasso Sea, and more recently, of other surface marine locations revealed that proton pump rhodopsin genes were extraordinarily diverse, suggesting that rhodopsin-based phototrophy is very important in marine ecosystems [35,41]. Bi-directional end-sequencing of large genomic clones provides information similar to that of random shotgun sequencing with the advantage that large genome fragments of single organisms remain available for study. This strategy has been used for the first comparative metagenomic studies in the water column [6] and in the deep sea [25]. Shotgun sequencing of low-diversity environments such as acid mines, symbiotic consortia or enrichments can result in construction of complete genome scaffolds and the portraying of whole organism metabolisms. The reconstruction of a putative reverse methanogenesis pathway in AOM archaea [16] or the complete genome scaffolds of *Leptospirillum* and *Ferroplasma* spp. [40] and of the versatile anammox bacteria *Kuenenia sttutgartensis* [39] have been made possible in this way.

Metagenomics is rapidly growing as sequencing costs decrease with the pyrosequencing and other new sequencing strategies. Hopefully, genomics of many more cultivated organisms will expand to provide necessary references, and single cell genomics, in which interesting progress is being made, will provide a valuable complementary approach.

3.3. Microbes and biogeochemical cycles

Efforts to link metabolic functions to organisms by taking into account the environmental context are paying off. New metabolisms are being discovered (AOM, anammox), as well as unexpected capabilities in lineages for which other metabolisms were thought to be general (phototrophic acidobacteria and euryarchaeota, ammonium oxidizing archaea). As some of these new players are abundant and active in soils or oceans, they have a substantial impact upon biogeochemical cycling at the planet scale, affecting in particular carbon and nitrogen cycles.

The major transformations of our N cycle comprehension affect nitrification ($NH_3 \rightarrow NO_2^- \rightarrow NO_3^-$), which was previously thought to be carried out by ammonia- and nitrite-

oxidizing bacteria, respectively. The discovery of anammox leading to the production of N_2 using NO_3^- has broken the linearity of the process, increasing the complexity of the cycle. At a global level, anammox is important in ocean sediments, but its actual contribution to the loss of fixed N in oceans is not yet known. The discovery of abundant aerobic ammonium oxidizing archaea (AOA) in soils and the deep ocean not only brings in another player, but also suggests that archaea, and not bacteria, are the ecologically relevant nitrifiers at a global level. AOA also represent a new connection with the C cycle, as they appear to be autotrophic [20,22]. Their abundance (up to one third of picoplankton biomass) in the deep ocean, where only residual heterotrophy sustained on sedimenting organic particles was thought to occur, is raising new questions about deep oceanic ecology [6,20,25]. One of the most significant findings for the C cycle is the importance of mixed metabolic strategies in the ocean. Photoheterotrophy using rhodopsin proton pumps in many different bacteria and even archaea appears broadly and abundantly distributed in photic regions [1,41], as well as that of aerobic anoxygenic phototrophy [2]. Lithoheterotrophy based on carbon monoxide oxidation also appears widespread in heterotrophic bacteria of the photic zone, and is responsible (to an unknown extent) for recirculation of greenhouse CO gas in the system [26]. Interestingly, CO oxidation appears to be far more important in the deep sea [25], where energy sources are limited. Versatile heterotrophs appear to adapt to oligotrophic ecosystems by exploiting any extra energy-generating metabolic reactions at their disposal.

4. Learning about (microbial) evolution

Biological evolution has a genetic basis. Therefore, knowledge of the true extent of phylogenetic diversity in living beings provides meaningful starting data for studying underlying patterns and processes of variation and for unraveling the history of major organismal lineages.

4.1. Macroevolution

Although the genetic diversity of multicellular eukaryotic organisms (animals, green plants, red and brown algae, fungi) seems large, it is rather modest when compared with the gigantic genetic diversity, partly fed by environmental molecular and genomic studies, observed at the eukaryotic domain and tree of life levels (Fig. 1). It turns out that most phylogenetic diversity is found in the microbial world, which had about three billion years more to evolve and diversify than multicellular taxa. How did the three domains of life evolve? What were the major evolutionary driving forces behind this? Except for the origin of life itself, few (if any) questions in biology are more unsettled and controversial than how bacterial and archaeal lines split from a common ancestor, and how eukaryotes originated. Mapping key molecular biology and metabolic features not affected by extensive horizontal gene transfer onto as many high taxa as possible is a feasible way of obtaining essential information that could help to reconstruct plausible scenarios concerning the evolution of domain or high-level taxa-specific traits.

From phylogenetic trees of genes and genomes, within-domain evolution reveals surprising trends that are not well understood. Bacterial phyla show an explosive radiation. It could be argued that this is the consequence of peculiarities of the prokaryotic mode and tempo of evolution. However, archaea, with their major split into two (or perhaps three) kingdoms, their more restricted diversity and their scaled differentiation of groups (e.g. hyperthermophiles at the base, methanogens and the like in the middle, and hyperhalophiles at the tip of the euryarchaeotal branch) behave essentially differently (Fig. 1). Why is this so? The bacterial aptitude to exploit any imaginable niche (the ‘‘microbial infallibility’’ principle proposed in the late 1940s) would explain why, with the exception of a few extreme environments dominated by archaea such as salt crystallizers or hydrothermal vents, bacterial diversity is several orders of magnitude larger. In contrast, archaea might be better adapted to energy-challenging conditions than the seemingly more versatile and opportunistic bacteria.

The case of the eukaryotic domain is no less problematic. The last fifteen years have seen the rise and demise of various phylogenetic hypotheses attempting to explain the evolution of eukaryotic taxa, the most famous of which was the archaezoa hypothesis postulating that true amitochondriate eukaryotes were basal in the eukaryotic branch of the tree. Heterogeneity of evolutionary rate, low and biased taxonomic sampling and simplistic models of sequence evolution partly accounted for the reconstruction of such erroneous phylogenies [32]. However, even with much better sampling and many phylogenetic markers sequenced for different cultivated groups arising from whole genome sequences or from expressed sequence tags (ESTs), most of the deepest nodes of the eukaryotic tree remain unresolved. A complementary approach to the inclusion of more genes for building eukaryotic trees is increasing taxonomic sampling, as many important groups of eukaryotes are not available in culture [28]. Improving models and tools of phylogenetic reconstruction, mapping the distribution of important characters on trees and broadening of the taxonomic sampling will be of great help.

4.2. Microevolution

The vast genetic diversity observed for prokaryotes and, more recently, microbial eukaryotes is raising important questions about speciation and also the mechanisms causing it at the finest population genetics level (see, for instance, the special issue ‘‘Species and Speciation in Microorganisms’’ published in 2006 by the Royal Society B). Although there is no consensual species concept for microbes, and sequence-based OTUs are used in practice, we can begin to make sense of massive accumulated data by prospecting patterns of genotypic and associated phenotypic variation to determine whether a coherent species concept can emerge and, most importantly, to explore the mechanisms involved in biological diversification. This is a general question raised by evolutionary biologists since Darwin’s time, who have built a body of

evolutionary ecology (macroecology) theory. Taking advantage of this knowledge, still largely unfamiliar to microbiologists, for building models and testing evolutionary hypotheses using microbial systems holds much promise. Not only will microbiologists improve their understanding of speciation mechanisms, but also, evolutionary ecologists will likely benefit from the use of microbial models because microbes usually have large population sizes and evolve more rapidly. A recent example was the study of the role of predation and immigration in adaptive radiation using a bacterial (prey)–protist (predator) system [12]. In addition to large population size, there is another fundamental difference in microbial systems, namely, the occurrence of variable rates of homologous recombination and potential high horizontal gene transfer levels. However, the role of recombination and clonal divergence in microbial speciation is not well understood [11], although homologous recombination is fairly common in bacterial populations, as attested by a variety of comparative genomic, metagenomic and multilocus sequence typing (MLST) analyses (e.g. [24,31,35]).

Besides the impact of recombination barriers in speciation, that of ecological versus geographic barriers remains another highly debated and unsolved issue. Traditionally, speciation is considered to occur in allopatry and, for some authors, prokaryotes and small protists, having large population sizes and no barriers to dispersal, are ubiquitous [9]. Consequently, the number of microbial species should be small. This is at odds with the immense microbial (prokaryotic and eukaryotic) genetic diversity observed in nature. Part of the answer to this paradox comes from the fact that much genetic variation is hidden (cryptic) under identical morphologies. This was naturally accepted from ancient times for bacteria, and sequence-based OTUs have replaced any morphological species concept, but it is not yet widely accepted by some traditional protistologists despite increasing evidence of huge amounts of genetic variation evidenced by MLST within single morphospecies [37]. Within-morphospecies lineages may be considered cryptic species (in some cases they do not even mate). They sometimes co-exist in the same environment, suggesting that speciation in sympatry can occur. An interesting possibility is that ecological barriers define ecotypes that become genetically isolated, forming new species. Studies of light-intensity-adapted *Prochlorococcus* genomic and ecological patterns in the ocean are providing insights concerning this question [10]. However, to evaluate the contribution of this mechanism to speciation, many more data are needed from different prokaryotic and protist models, as well as the potential selective pressures that may be operating taking into account prevalent environmental conditions, including data about local environmental heterogeneity.

Acknowledgments

We thank the Centre National de la Recherche Scientifique (CNRS) for support. We apologize to our microbiologist and protistologist colleagues for omitting many important original references in the field due to space limitations.

References

[1] Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L., Spudich, E.N., DeLong, E.F. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science 289, 1902–1906.

[2] Beja, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T., Eisen, J.A., Fraser, C.M., DeLong, E.F. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. Nature 415, 630–633.

[3] Boetius, A., Ravenschlag, K., Schubert, C.J., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Jorgensen, B.B., Witte, U., Pfannkuche, O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. Nature 407, 623–626.

[4] Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., Forterre, P. (2005) Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? Genome Biol. 6, R42.

[5] Bryant, D.A., Costas, A.M., Maresca, J.A., Chew, A.G., Klatt, C.G., Bateson, M.M., Tallon, L.J., Hostetler, J., Nelson, W.C., Heidelberg, J.F., Ward, D.M. (2007) Candidatus *Chloracidobacterium thermophilum*: an aerobic phototrophic Acidobacterium. Science 317, 523–526.

[6] DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., Martinez, A., Sullivan, M.B., Edwards, R., Brito, B.R., Chisholm, S.W., Karl, D.M. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. Science 311, 496–503.

[7] Dumont, M.G., Murrell, J.C. (2005) Stable isotope probing - linking microbial identity to function. Nat. Rev. Microbiol. 3, 499–504.

[8] Epstein, A., López-García, P. (online early) "Missing" protists: a molecular prospective, Biodiv. Conserv. doi: 10.1007/s10531-007-9250-y.

[9] Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. Science 296, 1061–1063.

[10] Follows, M.J., Dutkiewicz, S., Grant, S., Chisholm, S.W. (2007) Emergent biogeography of microbial communities in a model ocean. Science 315, 1843–1846.

[11] Fraser, C., Hanage, W.P., Spratt, B.G. (2007) Recombination and the nature of bacterial speciation. Science 315, 476–480.

[12] Fukami, T., Beaumont, H.J., Zhang, X.X., Rainey, P.B. (2007) Immigration history controls diversification in experimental adaptive radiation. Nature 446, 436–439.

[13] Gans, J., Wolinsky, M., Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 309, 1387–1390.

[14] Gerard, E., Guyot, F., Philippot, P., Lopez-Garcia, P. (2005) Fluorescence in situ hybridisation coupled to ultra small immunogold detection to identify prokaryotic cells using transmission and scanning electron microscopy. J. Microbiol. Methods 63, 20–28.

[15] Giovannoni, S., Stingl, U. (2007) The importance of culturing bacterioplankton in the "omics" age. Nat. Rev. Microbiol. 5, 820–826.

[16] Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., DeLong, E.F. (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. Science 305, 1457–1462.

[17] Harada, A., Ohtsuka, S., Horiguchi, T. (2007) Species of the parasitic genus *Duboscquella* are members of the enigmatic Marine Alveolate Group I. Protist 158, 337–347.

[18] Huang, W.E., Stoecker, K., Griffiths, R., Newbold, L., Daims, H., Whiteley, A.S., Wagner, M. (2007) Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. Environ. Microbiol. 9, 1878–1889.

[19] Hugenholtz, P., Goebel, B.M., Pace, N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol 180, 4765–4774.
[20] Ingalls, A.E., Shah, S.R., Hansman, R.L., Aluwihare, L.I., Santos, G.M., Druffel, E.R., Pearson, A. (2006) Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. Proc. Natl. Acad. Sci. USA 103, 6442–6447.
[21] Karner, M.B., DeLong, E.F., Karl, D.M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature 409, 507–510.
[22] Konneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437, 543–546.
[23] Kuypers, M.M., Sliekers, A.O., Lavik, G., Schmid, M., Jorgensen, B.B., Kuenen, J.G., Sinninghe Damste, J.S., Strous, M., Jetten, M.S. (2003) Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. Nature 422, 608–611.
[24] Lo, I., Denef, V.J., Verberkmoes, N.C., Shah, M.B., Goltsman, D., DiBartolo, G., Tyson, G.W., Allen, E.E., Ram, R.J., Detter, J.C., Richardson, P., Thelen, M.P., Hettich, R.L., Banfield, J.F. (2007) Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature 446, 537–541.
[25] Martin-Cuadrado, A.B., Lopez-Garcia, P., Alba, J.C., Moreira, D., Monticelli, L., Strittmatter, A., Gottschalk, G., Rodriguez-Valera, F. (2007) Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. PLoS ONE 2, e914.
[26] Moran, M.A., Miller, W.L. (2007) Resourceful heterotrophs make the most of light in the coastal ocean. Nat. Rev. Microbiol. 5, 792–800.
[27] Moreira, D., López-García, P. (2002) Molecular ecology of microbial eukaryotes unveils a hidden world. Trends Microbiol. 10, 31–38.
[28] Moreira, D., von der Heyden, S., Bass, D., Lopez-Garcia, P., Chao, E., Cavalier-Smith, T. (2007) Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. Mol. Phylogenet. Evol. 44, 255–266.
[29] Orphan, V.J., House, C.H., Hinrichs, K.U., McKeegan, K.D., DeLong, E.F. (2001) Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. Science 293, 484–487.
[30] Ottesen, E.A., Hong, J.W., Quake, S.R., Leadbetter, J.R. (2006) Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. Science 314, 1464–1467.
[31] Papke, R.T., Zhaxybayeva, O., Feil, E.J., Sommerfeld, K., Muise, D., Doolittle, W.F. (2007) Searching for species in haloarchaea. Proc. Natl. Acad. Sci. USA 104, 14092–14097.
[32] Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Muller, M., Le Guyader, H. (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc. R. Soc. Lond. B Biol. Sci. 267, 1213–1221.
[33] Reysenbach, A.L., Liu, Y., Banta, A.B., Beveridge, T.J., Kirshtein, J.D., Schouten, S., Tivey, M.K., Von Damm, K.L., Voytek, M.A. (2006) A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. Nature 442, 444–447.
[34] Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1, 283–290.
[35] Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.H., Falcon, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M., Venter, J.C. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. 5, e77.
[36] Schleper, C., Jurgens, G., Jonuscheit, M. (2005) Genomic studies of uncultivated archaea. Nat. Rev. Microbiol. 3, 479–488.
[37] Slapeta, J., Lopez-Garcia, P., Moreira, D. (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. Mol. Biol. Evol. 23, 23–29.
[38] Stepanauskas, R., Sieracki, M.E. (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc. Natl. Acad. Sci. USA 104, 9052–9057.
[39] Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M.W., Horn, M., Daims, H., Bartol-Mavel, D., Wincker, P., Barbe, V., Fonknechten, N., Vallenet, D., Segurens, B., Schenowitz-Truong, C., Medigue, C., Collingro, A., Snel, B., Dutilh, B.E., Op den Camp, H.J., van der Drift, C., Cirpus, I., van de Pas-Schoonen, K.T., Harhangi, H.R., van Niftrik, L., Schmid, M., Keltjens, J., van de Vossenberg, J., Kartal, B., Meier, H., Frishman, D., Huynen, M.A., Mewes, H.W., Weissenbach, J., Jetten, M.S., Wagner, M., Le Paslier, D. (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature 440, 790–794.
[40] Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43.
[41] Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., Smith, H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.
[42] Wagner, M., Nielsen, P.H., Loy, A., Nielsen, J.L., Daims, H. (2006) Linking microbial community structure with function: fluorescence in situ hybridization-microautoradiography and isotope arrays. Curr. Opin. Biotechnol 17, 83–91.

Purificación López-García*
David Moreira
Unité d'Ecologie, Systématique et Evolution, UMR CNRS 8079, Université Paris-Sud, bâtiment 360, 91405 Orsay Cedex, France
*Corresponding author.
E-mail address: puri.lopez@u-psud.fr (P. López-García)

16 November 2007

Available online 14 December 2007